

RNAstructure: web servers for RNA secondary structure prediction and analysis

Stanislav Bellaousov^{1,2}, Jessica S. Reuter^{1,2}, Matthew G. Seetin^{1,2} and David H. Mathews^{1,2,3,*}

¹Department of Biochemistry and Biophysics, University of Rochester Medical Center, 601 Elmwood Avenue, Box 712, Rochester, NY 14642, USA, ²Center for RNA Biology, University of Rochester Medical Center, 601 Elmwood Avenue, Box 712, Rochester, NY 14642, USA and ³Department of Biostatistics and Computational Biology, University of Rochester Medical Center, 601 Elmwood Avenue, Box 712, Rochester, NY 14642, USA

Received January 30, 2013; Revised March 22, 2013; Accepted March 31, 2013

ABSTRACT

RNAstructure is a software package for RNA secondary structure prediction and analysis. This contribution describes a new set of web servers to provide its functionality. The web server offers RNA secondary structure prediction, including free energy minimization, maximum expected accuracy structure prediction and pseudoknot prediction. Bimolecular secondary structure prediction is also provided. Additionally, the server can predict secondary structures conserved in either two homologs or more than two homologs. Folding free energy changes can be predicted for a given RNA structure using nearest neighbor rules. Secondary structures can be compared using circular plots or the scoring methods, sensitivity and positive predictive value. Additionally, structure drawings can be rendered as SVG, postscript, jpeg or pdf. The web server is freely available for public use at: <http://rna.urmc.rochester.edu/RNAstructureWeb>.

INTRODUCTION

RNA is an important biomolecule, as a catalyst (1), a director of post-transcriptional modification (2) and gene regulation (3), a target of drugs (4,5) and also a pharmaceutical (6,7). Its structure is generally hierarchical (8). Primary structure is the sequence of nucleotides, and is a covalent bond structure. Secondary structure is the set of the canonical base pairs, and tertiary structure is the set of additional contacts and the complete three-dimensional structure.

Because RNA folding is generally hierarchical, secondary structure can often be predicted and analyzed without predicting tertiary structure. Secondary structure

prediction can provide a framework for understanding the mechanism of action of RNA. Secondary structure prediction is also an important consideration in the design of siRNA and antisense DNA oligonucleotides both to avoid structure in the mRNA target of these agents and in avoiding self-structure in agents, either of which prevents them from hybridizing with their targets (9–14). RNA structure prediction can also make predictions about which regions of sequence are accessible for interacting with proteins (15). Finally, secondary structure prediction can be used to identify novel functional RNA sequences encoded in genomes (16–18).

RNAstructure is a software package for RNA secondary structure prediction and analysis (19). It was first reported in 1998 as a free energy minimization program for Microsoft Windows (20). It has been expanded to include methods for predicting bimolecular structure (12), conserved structures in multiple homologs (21–24) and siRNA design (9). Several methods are available for predicting structures for a single sequence, including maximum expected accuracy (25), stochastic sampling (26), exhaustive traceback (27) and pseudoknot prediction (28). Graphical user interfaces are provided for Microsoft Windows, Macintosh OS-X and Linux. Command line interfaces are also available for these operating systems. Finally, the set of underlying C++ classes are available as a library for use by programmers.

This report describes a new set of web servers that provide the functionality of RNAstructure. These web servers are open to the public and can be found at <http://rna.urmc.rochester.edu/RNAstructureWeb>.

ORGANIZATION

The RNAstructure web servers are organized around two schemes. The first scheme provides specific programs for web users. Table 1 shows a list of programs that are available. The second scheme is a set of themes defined by the

*To whom correspondence should be addressed. Tel: +1 585 275 1734; Fax: +1 585 275 6007; Email: david_mathews@urmc.rochester.edu

Table 1. List of individual programs provided by the RNAstructure web server

Servers	References	Description
<i>AllSub</i>	(27,29)	Generate all possible low free energy structures for a nucleic acid sequence.
<i>bifold</i>	(12)	Predict the lowest free energy structure for two interacting sequences, allowing intramolecular base pairs.
<i>bipartition</i>	(30)	Perform a partition function calculation for two interacting nucleic acid sequences. Intramolecular pairs are not allowed.
<i>CircleCompare</i>	(19)	Compare two structures using overlaid circular plots to emphasize pairing differences and pseudoknots.
<i>ct2dot</i>	(19)	Convert a CT-formatted structure into a dot bracket file.
<i>dot2ct</i>	(19)	Convert a dot bracket file into a CT file.
<i>draw</i>	(19)	Draw the secondary structure of a nucleic acid strand, with or without color annotation.
<i>DuplexFold</i>	(31)	Predict the lowest free energy structure for two interacting sequences, not allowing intramolecular base pairs.
<i>Dynalign</i>	(17,21,32)	Calculate the lowest free energy secondary structures common to two unaligned sequences.
<i>efn2</i>	(33)	Calculate the folding free energy of structures in a CT file.
<i>EnsembleEnergy</i>	(30)	Calculate the ensemble folding free energy change for a sequence.
<i>Fold</i>	(33,34)	Predict the lowest free energy structure and a set of low free energy structures for a sequence.
<i>MaxExpect</i>	(25)	Generate a structure or structures composed of highly probable base pairs.
<i>Multilign</i>	(22)	Predict low free energy secondary structures common to three or more sequences using progressive iterations of Dynalign.
<i>oligoscreen</i>	(35)	Predict stability of hybridization and self-structure for a list of oligonucleotides.
<i>OligoWalk for siRNA Design</i>	(9,11)	Design efficient siRNAs for a given miRNA target.
<i>partition</i>	(30)	Perform a partition function calculation on a single sequence to calculate base pair probabilities.
<i>PARTS</i>	(23,36)	Predict the common secondary structure, including base pair probabilities, for two unaligned sequences.
<i>ProbablePair</i>	(30)	Generate secondary structures composed of base pairs with probabilities that exceed a specified threshold.
<i>ProbKnot</i>	(28)	Predict a secondary structure composed of probable base pairs, which might include pseudoknots.
<i>RemovePseudoknots</i>	(19,37)	Remove pseudoknots from a structure.
<i>scorer</i>	(31,30)	Calculate the sensitivity and positive predictive value (PPV) for a predicted as compared to the accepted structure.
<i>stochastic</i>	(26,36)	Generate a representative sample of structures using stochastic sampling.
<i>TurboFold/TurboKnot</i>	(24,38)	Calculate the conserved structures of three or more unaligned sequences using iteratively refined partition functions.

biological problem. For example, to predict a secondary structure for a single sequence, the server called *Predict a Secondary Structure* can be used to run calculations by multiple programs. The list of themes and the exact programs included are listed in Table 2. This approach is designed to be user-friendly because it bundles all the available methods for addressing a specific problem, and for this reason most users will want to use these servers.

INPUT

The structure prediction servers work using either bare sequences pasted into a window, or FASTA-formatted files uploaded to the server. The exception to this is the set of servers that predict a structure common to three or more sequences. For these servers, including *Multilign* and *TurboFold*, the sequence window requires multiple FASTA-formatted sequences. Alternatively, a FASTA-formatted file with multiple sequences can be uploaded.

Servers that act on structures, such as *CircleCompare*, *draw*, *efn2*, *RemovePseudoknots* or *scorer*, require an upload of a CT file. These can be generated manually or by a structure prediction server. Another alternative is to use a dot-bracket formatted structure and convert it to CT format using the *dot2ct* component of the RNAstructure server.

For each individual server, sample input data can be generated automatically to illustrate format and to provide a test case for the server. For structure prediction

servers, clicking a link pastes sample sequences into the sequence windows. For the servers that act on structures, a link is provided to download a sample CT file, and this can be uploaded back to the server as sample data.

For *Fold*, *partition* and *Predict a Secondary Structure*, structure prediction can be restrained using SHAPE mapping data, when available, to improve the accuracy of structure prediction (39). *AllSub*, *Dynalign*, *Fold*, *partition* and *Predict a Secondary Structure* accept folding constraints, including the ability for force specific base pairs, forbid specific pairs, force a nucleotide paired, force a nucleotide unpaired, and specify that a nucleotide is accessible to chemical modification (33,34). These constraints and restraints are mediated by file uploads.

OUTPUT FORMATS

Once a calculation is submitted, a notification page is displayed. This page continues to refresh until the calculation output is available. At that time, the page is replaced with a page that contains the output. This page can be bookmarked and returned to at a later time. Alternatively, an e-mail address can be provided as part of the input data. If this is done, an e-mail is sent to the address when the calculation is complete and the results are available.

The structure prediction servers display the predicted structure using an SVG drawing, which can be rendered by web browsers with an SVG plugin (Figure 1). If more than one structure is predicted, i.e. suboptimal structures

Table 2. List of RNAstructure web servers organized around a problem

Servers	Programs run (see Table 1 for description)
Predict a secondary structure	<i>Fold, MaxExpect, ProbKnot, partition</i>
Predict a secondary structure common to two sequences	<i>Dynalign, Parts</i>
Predict a secondary structure common to three or more sequences	<i>Multalign, TurboFold/TurboKnot</i>
Predict a bimolecular secondary structure	<i>bifold, DuplexFold</i>

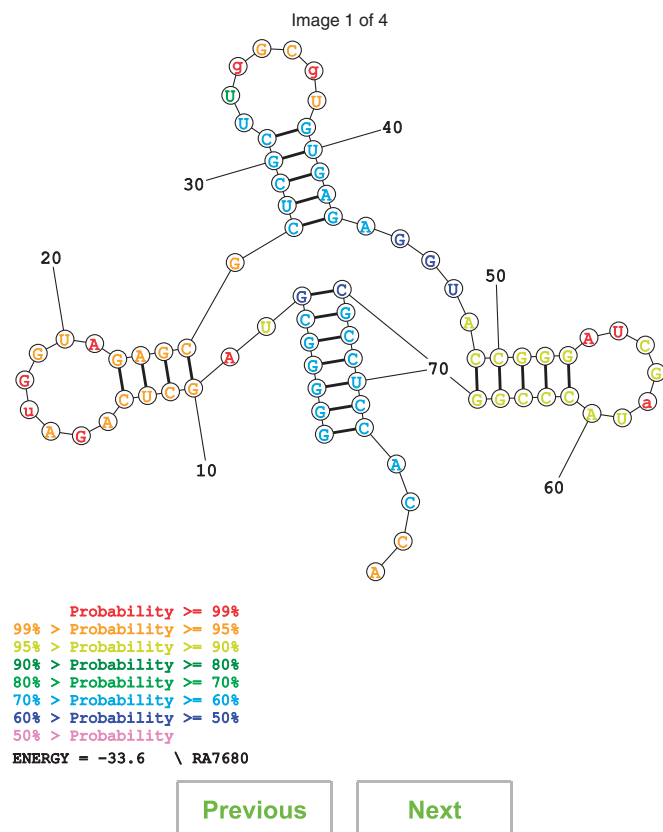


Figure 1. Sample structure prediction. This is the predicted lowest free energy structure for the sample sequence for the 'Predict a Secondary Structure' server. This sequence is the tRNA RA7680 from the Sprinzl database (41). It is color annotated according to base pairing probability; more probable pairs are more likely to be correctly predicted than less probable pairs (30).

or a structure sample is present, 'previous' and 'next' buttons are displayed to enable scrolling between structures. For each predicted structure, the structure can be downloaded as a jpeg, svg, pdf, postscript or CT file.

Each output page displays the RNAstructure executable and the command line used to generate the results (40). Each input file is a link, allowing download of the processed-form data. This makes clear exactly what calculation is performed, and facilitates a user learning the capabilities of the RNAstructure software.

IMPLEMENTATION

The web servers use PHP to process queries. The server does not rely on JavaScript running on the client side,

although a single exception to this is the previously available *OligoWalk* web server (11). The server itself is a computer cluster with a head node and five compute nodes, each with two dual core processors. A queuing system prioritizes the order at which queries are executed, which is generally in the order in which queries were made. The exceptions to this rule are calculations submitted for *Dynalign*, *Multalign* or *TurboFold*, which are submitted to the server as multiple-core calculations. These may need to wait in the queue at times when a number of other time-consuming calculations are already running.

The RNAstructure web servers have limitations on the calculations that can be done. This ensures that the resource will be available to the broader community with reasonable wait times. A list of limitations is available with the online help. Currently, for example, the single sequence structure prediction methods are limited to sequences of 2500 nucleotides or less. *TurboFold* is limited to a maximum of 12 sequences up to 1600 nucleotides in length. If the required calculation exceeds the web server requirements, the software can be downloaded (<http://rna.urmc.rochester.edu/RNAstructure.html>) and run locally.

HELP

Extensive online help is available for using the RNAstructure web server. The help documents each of the input and output fields, and is organized by individual server. For each server help page, a link is provided to the underlying RNAstructure executable (19), which provides additional details about the program. A separate help page details the server limitations, as explained above.

CONCLUSION

RNAstructure is designed to be a user-friendly software package, accessible to the community of investigators studying RNA (19). These new web servers continue this tradition, and provide the software to a wider user base. The accuracy of the algorithms has been extensively tested in prior publications. As the algorithms in RNAstructure are improved, and new algorithms developed, the web servers will be updated to continue to make RNAstructure available to the community.

FUNDING

National Institutes of Health (NIH) [R01 GM076485 to D.H.M.]. Funding for open access charge: NIH [R01 GM076485].

Conflict of interest statement. None declared.

REFERENCES

- Doudna, J.A. and Cech, T.R. (2002) The chemical repertoire of natural ribozymes. *Nature*, **418**, 222–228.
- Bachellerie, J.P., Cavaille, J. and Huttenhofer, A. (2002) The expanding snoRNA world. *Biochimie*, **84**, 775–790.
- Gong, C. and Maquat, L.E. (2011) lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature*, **470**, 284–288.
- Suchack, S.J. and Wong, C.H. (2000) RNA as a target for small molecules. *Curr. Opin. Chem. Biol.*, **4**, 678–686.
- Guan, L. and Disney, M.D. (2012) Recent advances in developing small molecules targeting RNA. *ACS Chem. Biol.*, **7**, 73–86.
- Siomi, H. and Siomi, M.C. (2009) On the road to reading the RNA-interference code. *Nature*, **457**, 396–404.
- Castanotto, D. and Rossi, J.J. (2009) The promises and pitfalls of RNA-interference-based therapeutics. *Nature*, **457**, 426–433.
- Tinoco, I. Jr and Bustamante, C. (1999) How RNA folds. *J. Mol. Biol.*, **293**, 271–281.
- Lu, Z.J. and Mathews, D.H. (2007) Efficient siRNA selection using hybridization thermodynamics. *Nucleic Acids Res.*, **36**, 640–647.
- Lu, Z.J. and Mathews, D.H. (2008) Fundamental differences in the equilibrium considerations for siRNA and antisense oligodeoxynucleotide design. *Nucleic Acids Res.*, **36**, 3738–3745.
- Lu, Z.J. and Mathews, D.H. (2008) OligoWalk: an online siRNA design tool utilizing hybridization thermodynamics. *Nucleic Acids Res.*, **36**, W104–W108.
- Mathews, D.H., Burkard, M.E., Freier, S.M., Wyatt, J.R. and Turner, D.H. (1999) Predicting oligonucleotide affinity to nucleic acid targets. *RNA*, **5**, 1458–1469.
- Tafer, H., Ameres, S.L., Obernosterer, G., Gebeshuber, C.A., Schroeder, R., Martinez, J. and Hofacker, I.L. (2008) The impact of target site accessibility on the design of effective siRNAs. *Nat. Biotechnol.*, **26**, 578–583.
- Long, D., Lee, R., Williams, P., Chan, C.Y., Ambros, V. and Ding, Y. (2007) Potent effect of target structure on microRNA function. *Nat. Struct. Mol. Biol.*, **14**, 287–294.
- Li, X., Quon, G., Lipshitz, H.D. and Morris, Q. (2010) Predicting in vivo binding sites of RNA-binding proteins using mRNA secondary structure. *RNA*, **16**, 1096–1107.
- Washietl, S., Hofacker, I.L. and Stadler, P.F. (2005) Fast and reliable prediction of noncoding RNAs. *Proc. Natl Acad. Sci. USA*, **102**, 2454–2459.
- Uzilov, A.V., Keegan, J.M. and Mathews, D.H. (2006) Detection of non-coding RNAs on the basis of predicted secondary structure formation free energy change. *BMC Bioinformatics*, **7**, 173.
- Torarinsson, E., Sawera, M., Havgaard, J.H., Fredholm, M. and Gorodkin, J. (2006) Thousands of corresponding human and mouse genomic regions unalignable in primary sequence contain common RNA structure. *Genome Res.*, **16**, 885–889.
- Reuter, J.S. and Mathews, D.H. (2010) RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*, **11**, 129.
- Mathews, D.H., Andre, I.C., Kim, J., Turner, D.H. and Zuker, M. (1998) An updated recursive algorithm for RNA secondary structure prediction with improved thermodynamic parameters. In: Leontis, N.B. and SantaLucia, J. Jr (eds), *Molecular Modeling of Nucleic Acids*. American Chemical Society, Washington, D.C., pp. 246–257.
- Mathews, D.H. and Turner, D.H. (2002) Dynalign: an algorithm for finding the secondary structure common to two RNA sequences. *J. Mol. Biol.*, **317**, 191–203.
- Xu, Z. and Mathews, D.H. (2011) Multalign: an algorithm to predict secondary structures conserved in multiple RNA sequences. *Bioinformatics*, **27**, 626–632.
- Harmanci, A.O., Sharma, G. and Mathews, D.H. (2008) PARTS: probabilistic alignment for RNA joint secondary structure prediction. *Nucleic Acids Res.*, **36**, 2406–2417.
- Harmanci, A.O., Sharma, G. and Mathews, D.H. (2011) TurboFold: iterative probabilistic estimation of secondary structures for multiple RNA sequences. *BMC Bioinformatics*, **12**, 108.
- Lu, Z.J., Gloor, J.W. and Mathews, D.H. (2009) Improved RNA secondary structure prediction by maximizing expected pair accuracy. *RNA*, **15**, 1805–1813.
- Ding, Y. and Lawrence, C.E. (2003) A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic Acids Res.*, **31**, 7280–7301.
- Duan, S., Mathews, D.H. and Turner, D.H. (2006) Interpreting oligonucleotide microarray data to determine RNA secondary structure: application to the 3' end of Bombyx mori R2 RNA. *Biochemistry*, **45**, 9819–9832.
- Bellaousov, S. and Mathews, D.H. (2010) ProbKnot: fast prediction of RNA secondary structure including pseudoknots. *RNA*, **16**, 1870–1880.
- Wuchty, S., Fontana, W., Hofacker, I.L. and Schuster, P. (1999) Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers*, **49**, 145–165.
- Mathews, D.H. (2004) Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization. *RNA*, **10**, 1178–1190.
- Piekna-Przybylska, D., DiChiacchio, L., Mathews, D.H. and Bambara, R.A. (2009) A sequence similar to tRNA^{3Lys} gene is embedded in HIV-1 U3/R and promotes minus strand transfer. *Nat. Struct. Mol. Biol.*, **17**, 83–89.
- Harmanci, A.O., Sharma, G. and Mathews, D.H. (2007) Efficient pairwise RNA structure prediction using probabilistic alignment constraints in dynalign. *BMC Bioinformatics*, **8**, 130.
- Mathews, D.H., Sabina, J., Zuker, M. and Turner, D.H. (1999) Expanded sequence dependence of thermodynamic parameters provides improved prediction of RNA secondary structure. *J. Mol. Biol.*, **288**, 911–940.
- Mathews, D.H., Disney, M.D., Childs, J.L., Schroeder, S.J., Zuker, M. and Turner, D.H. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl Acad. Sci. USA*, **101**, 7287–7292.
- Matveeva, O.V., Mathews, D.H., Tsodikov, A.D., Shabalina, S.A., Gesteland, R.F., Atkins, J.F. and Freier, S.M. (2003) Thermodynamic criteria for high hit rate antisense oligonucleotide design. *Nucleic Acids Res.*, **31**, 4989–4994.
- Harmanci, A.O., Sharma, G. and Mathews, D.H. (2009) Stochastic sampling of the RNA structural alignment space. *Nucleic Acids Res.*, **37**, 4063–4075.
- Smit, S., Rother, K., Heringa, J. and Knight, R. (2008) From knotted to nested RNA structures: a variety of computational methods for pseudoknot removal. *RNA*, **14**, 410–416.
- Seetin, M.G. and Mathews, D.H. (2012) TurboKnot: rapid prediction of conserved RNA secondary structures including pseudoknots. *Bioinformatics*, **28**, 792–798.
- Deigan, K.E., Li, T.W., Mathews, D.H. and Weeks, K.M. (2009) Accurate SHAPE-directed RNA structure determination. *Proc. Natl Acad. Sci. USA*, **106**, 97–102.
- Gruber, A.R., Lorenz, R., Bernhart, S.H., Neubock, R. and Hofacker, I.L. (2008) The Vienna RNA websuite. *Nucleic Acids Res.*, **36**, W70–W74.
- Sprinzl, M. and Vassilenko, K.S. (2005) Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.*, **33**, D139–D140.